2/10/22:

Today: computing coarse correlated equilibria with "natural" dynamics: online learning!

Def: Let σ be a distribution over $S = S_1 \times S_2 \times \dots \times S_k$.

Then σ is a coarse correlated equilibrium if

$$\mathop{E}_{s\sim\sigma}[c_i(s)] \leq \mathop{E}_{s\sim\sigma}[c_i(s_{-i}, s_i')] \quad \forall i \in [k], \ \forall s_i' \in S_i$$

Detour to online learning:

Multiarmed bandits:

- Action set A (arms).
- At time $t = 1, 2, \dots, T$:
  - Algorithm picks distribution p^t over A
  - Adversary picks cost vector $c^t: A \to [0,1]$
  - Action $a^t \sim p^t$, algorithm incurs cost $c^t(a^t)$
  - Algorithm learns either $c^t(a^t)$ (bandit setting) or $c^t(a)$ $\forall a \in A$ (experts setting)

Goal for algorithm: minimize total cost!

Need to be a little more careful about adversary

Def: An adaptive adversary takes as input

1) Algorithm A, 2) time t,

3) distributions $p^1, p^2, \ldots, p^t$ produced by A

4) Realized actions $a^1, a^2, \ldots, a^{t-1}$ from past

and outputs costs $c^t : A \to [0,1]$

Def: An oblivious (or non-adaptive) adversary is

is an adversary that depends only on A and t

Equivalent: fixes cost functions at beginning, knowing A.

Surprising: not much difference in what can be achieved!

For us: - care about adaptive adversaries

- mostly analyze oblivious for simplicity

- details in textbooks

Thm: No algorithm can be competitive with the best action sequence in hindsight (adaptive adversary)

Pf: Let $|A| = 2$

Adversary: If $p^t(0) \geq \frac{1}{2}$: $c^t(0) = 1$, $c^t(1) = 0$

If $p^t(1) > \frac{1}{2}$: $c^t(0) = 0$, $c^t(1) = 1$

$\Rightarrow$ at every time t, algorithm has expected cost $\geq \frac{1}{2}$

$\Rightarrow$ expected cost of algorithm is $\geq \frac{T}{2}$

But there is some sequence of actions with total cost $= 0$ ✓

New benchmark: not best action sequence, but best action

Def: The regret of an action sequence $a^1, a^2, a^3, \ldots, a^T$ with respect to $a \in A$ is

$$R_T(a) = \frac{1}{T}\left(\sum_{t=1}^{T} c^t(a^t) - \sum_{t=1}^{T} c^t(a) \right)$$

Def: If $\mathcal{A}$ is an online learning algorithm, then its expected regret at time T with respect to $a \in A$ is

$$E\left[R_T^{\mathcal{A}}(a)\right] = \frac{1}{T}\left(\sum_{t=1}^{T} \underset{a^t \sim p^t}{E}\left[c^t(a^t)\right] - \sum_{t=1}^{T} c^t(a)\right)$$

Def: Algorithm $\mathcal{A}$ is a no-regret algorithm (or has no-regret) if for every adversary, for every $a \in A$,

$$\lim_{T \to \infty} E\left[R_T^{A}(a)\right] = 0 \qquad \left(\text{or } E\left[R_T^{A}(a)\right] = o(1)\right)$$

Amazing fact: no-regret algorithms exist, and are pretty simple!

Relationship to game theory:

Can model playing a game as online learning!

- Particularly attractive if player does not know details of game

Suppose you're player i, play game T times.

Think of other players' mixed strategies as adversary!

At time t:

- every player $j \neq i$ chooses some mixed strategy p_j^t over S_j

- define $c_i^t(a) = \underset{\substack{a_j \sim p_j^t \\ \forall j \neq i}}{E}\left[C_i(a_1, a_2, \ldots, a_{i-1}, a, a_{i+1}, \ldots, a_n) \right]$

learning cost

game cost

$n = \#$ players

  - Possibly unknown to player!

- Could use a no-regret algorithm to choose p_i^t!

Note: adversary is adaptive!

Informal claim: "Rational" way for players to act!

Connection to equilibria:

Game with k players, cost functions $C_i : S \to [0,1]$

play T times

- Player i uses algorithm A_i
- Let p_i^t be mixed strategy used by player i at time t
- Let $\sigma^t = \prod_{i=1}^{k} p_i^t$ be product distribution over S induced by individual player distributions
- Let $\sigma = \frac{1}{T} \sum_{t=1}^{T} \sigma^t$ be "average" distribution.

  Interpretation: sample t uniformly from $[T]$, then sample from σ^t

  Note: Not product distribution!

- For each $i \in [k]$, $t \in [T]$, and $a \in S_i$, let

$$c_i^t(a) = \mathop{E}_{s \sim \sigma^t} \left[C_i(s_{-i}, a) \right]$$

Note:
$$\mathop{E}_{a \sim p_i^t} \left[c_i^t(a) \right] = \mathop{E}_{a \sim p_i^t} \left[\mathop{E}_{s \sim \sigma^t} \left[C_i(s_{-i}, a) \right] \right]$$
$$= \mathop{E}_{s \sim \sigma^t} \left[C_i(s) \right]$$

Thm: Sps that $\mathbb{E}[R_T^{A_i}(a)] \le \varepsilon \quad \forall i \in [k], \forall a \in S_i$.

Then σ is an ε-approximate coarse correlated equilibrium:

$$\mathop{\mathbb{E}}_{s \sim \sigma}[c_i(s)] \le \mathop{\mathbb{E}}_{s \sim \sigma}[c_i(s_{-i}, s_i')] + \varepsilon \quad \forall i \in [k], \forall s_i' \in S_i$$

Interpretation: average distribution converges to a CCE!

"empirical distribution of play" converges to a CCE

Pf:

$$\mathop{\mathbb{E}}_{s \sim \sigma}[c_i(s)] - \mathop{\mathbb{E}}_{s \sim \sigma}[c_i(s_{-i}, s_i')] = \qquad \text{(want to show } \le \varepsilon\text{)}$$

$$= \frac{1}{T} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{s \sim \sigma^t}[c_i(s)] - \frac{1}{T} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{s \sim \sigma^t}[c_i(s_{-i}, s_i')] \qquad \text{(def of } \sigma\text{)}$$

$$= \frac{1}{T}\left(\sum_{t=1}^{T} \mathop{\mathbb{E}}_{a^t \sim p_i^t}[c_i^t(a^t)] - \sum_{t=1}^{T} c_i^t(s_i') \right) \qquad \text{(def of } c_i^t\text{)}$$

$$= \mathbb{E}[R_T^{A_i}(s_i')] \qquad \text{(def of regret)}$$

$$\le \varepsilon \qquad \checkmark$$